ROBUSTNESS ASPECTS OF THE RASF

Katriina Halonen[1], *Pauli Kuosmanen*[2] *and Jaakko Astola*[3]

Tampere University of Technology, Signal Processing Laboratory
P.O.Box 553, FIN-33101 Tampere, Finland
E-mail:[1]kha@cs.tut.fi, [2]pqo@cs.tut.fi, [3]jta@cs.tut.fi

ABSTRACT

In this paper we study some robustness aspects of the Recursive Approaching Signal Filter (RASF) by deriving the influence functions for every iteration stage. The existence of the influence function for the whole estimator is studied. Guidelines for the selection of a suitable weighting parameter and initial reference values in order to support the robustness of the estimator are also given based on the influence functions. An example of a case where the behavior of the RASF is extremely non-robust in the influence function sense is presented.

1. INTRODUCTION

The RASF [5] is a recursive method for removing noise from signals, based on the idea that a filtered signal output is ideally closer to the unknown desired signal than the input. The larger the difference between the original input signal and the filtered output at a certain point, the more probably the sample at that point is an outlier and should be dealt with accordingly. A new filter can be constructed as a weighted average with each window sample weighted according to how far it is from the previous filtered output at the window center. The filter is made recursive by computing the weights in each filtering iteration in the above manner. These iterations should eventually converge to a fixed point.

The discrete version of the RASF is defined as follows, [5]: The output of the α-weighted discrete Recursive Approaching Signal Filter at x_0 using a window of length $2L+1$ is the limit of the sequence $R^{(k)}(x_0)$ as $k \to \infty$,

$$\begin{cases} W_k(x;x_0) = e^{-\alpha|R^{(k)}(x_0)-g(x)|} \\ R^{(k+1)}(x_0) = \dfrac{\sum_{t=x_0-L}^{x_0+L} W_k(t;x_0)g(t)}{\sum_{t=x_0-L}^{x_0+L} W_k(t;x_0)} \end{cases}, \qquad (1)$$

where the initial condition is $R^{(1)}(x_0) = T[g(x_0)]$, $g(\cdot)$ is the input signal, α is a positive real number, and $T(\cdot)$ is a filter, typically lowpass (e.g., median or mean). The function $W_k(x;x_0)$ contains the weight for the signal value at window position $x - x_0$ when the filtering window is centered at x_0 at the k^{th} iteration step.

The RASF was designed as a flexible filter which can be easily transformed into different filters, either linear or nonlinear, by changing the weighting parameter α [5]. The estimator was designed to preserve edges and other details while removing noise.

2. INFLUENCE FUNCTION

The *influence function* (IF) is a tool for inspecting some aspects of the robustness of an estimator. It measures the effect on the output of the estimator caused by infinitesimal contamination at a certain point in the model distribution, that is, slight deviation from the model, and it is defined as follows [1]:

$$\mathrm{IF}_T(x;F) = \lim_{t\to 0^+} \frac{T((1-t)F + t\Delta_x) - T(F)}{t} \qquad (2)$$

in those x where the limit exists. T is the estimator, F is the assumed distribution with the density $f(x)$, and Δ_x is the distribution with an impulse at point x. Fig. 1 shows the influence function of the mean and the median [1], assuming the standard normal distribution (or any other symmetric, zero-mean distribution, with changes only in the constant value of the IF of the median; in addition, the cumulative distribution function of F must be continuous at 0 for the median to exist).

Certain measures for the robustness of the estimator can be derived from the influence function, the most straightforward of which are the gross-error sensitivity, the rejection point, and the asymptotic variance [1]. The *gross-error sensitivity* can be defined as the supremum of the absolute of all values of the influence function, and it measures the worst influence which a small amount of contamination can have on the output of the estimator. The *rejection point* is the smallest positive

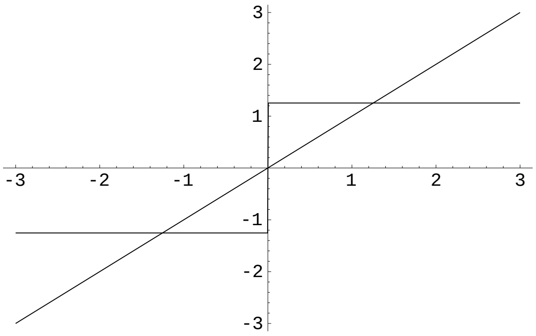


Figure 1: Influence function of the mean ($y = x$) and of the median ($y = \frac{sign(x)}{2f(0)}$).

x such that the influence function is zero outside the range $[-x, x]$, and, if finite, it indicates absolute rejection of some samples considered as outliers. For the mean and the median, the rejection point is ∞, that is, neither estimator totally discards outliers, but the gross-error sensitivity for the median is $\frac{1}{2f(0)}$ and ∞ for the mean, indicating better robustness for the median.

The *asymptotic variance* is defined as

$$V_T(F) = \int_{-\infty}^{\infty} (\mathrm{IF}_T(x; F))^2 dF(x), \tag{3}$$

where the notation $dF(x)$ is read as $f(x)dx$. In most cases the distribution of the estimator is asymptotically normal, i.e., as the window size $N = 2L + 1$ grows to infinity. The distribution of the estimator output T_N tends weakly to the normal distribution, having zero mean and the variance $V_T(F)/N$, assuming that $T(F) = 0$.

The influence function of the first iteration of the RASF for a symmetric zero mean distribution F can be shown to be as follows (the derivation is presented in the Appendix):

$$\mathrm{IF}_1(x; F) = \frac{xe^{-\alpha|x|} + \mathrm{IF}_T(x;F) \int_{-\infty}^{\infty} \alpha|\xi|e^{-\alpha|\xi|}dF(\xi)}{\int_{-\infty}^{\infty} e^{-\alpha|\xi|}dF(\xi)}, \tag{4}$$

where we use the subscript of IF to denote the number of the iteration of the RASF to shorten the notation. Elsewhere, the subscript denotes the estimator itself. The assumption in the derivation for Eq. 4 is that $T(F) = 0$.

From Eq. 4, we can further derive the influence functions for any iteration round, obtaining

$$\mathrm{IF}_{\mathrm{k}}(x; F) = \frac{xe^{-\alpha|x|} + \mathrm{IF}_{\mathrm{k\text{-}1}}(x;F) \int_{-\infty}^{\infty} \alpha|\xi|e^{-\alpha|\xi|}dF(\xi)}{\int_{-\infty}^{\infty} e^{-\alpha|\xi|}dF(\xi)}. \tag{5}$$

Let us denote $\int_{-\infty}^{\infty} \alpha|\xi|e^{-\alpha|\xi|}dF(\xi)$ in Eq. 5 with the symbol b and $\int_{-\infty}^{\infty} e^{-\alpha|\xi|}dF(\xi)$ with the symbol c, so that we can write the formula in the following form:

$$\mathrm{IF}_{\mathrm{k}}(x; F) = \frac{xe^{-\alpha|x|} + b \cdot \mathrm{IF}_{\mathrm{k\text{-}1}}(x; F)}{c}. \tag{6}$$

The influence function for the whole estimator is the limit of this function sequence as $k \to \infty$, if this limit exists. The convergence of the sequence is considered in the next section. If the sequence converges and the limit exists, we can write Eq. 6 in the form

$$\mathrm{IF}_{\infty}(x; F) = \frac{xe^{-\alpha|x|} + b \cdot \mathrm{IF}_{\infty}(x; F)}{c}. \tag{7}$$

and solve it for IF_{∞}. Then we obtain

$$\mathrm{IF}_{\infty}(x; F) = \frac{xe^{-\alpha|x|}}{c - b}. \tag{8}$$

The above reasoning is consistent with what we know from the theory of M-estimators [2]. It has been shown in [4] that the RASF is an M-estimator, whose defining ψ-function is $\psi(x) = xe^{-\alpha|x|}$, with the estimator converging to a limit. For M-estimators, the influence function is of the form

$$\mathrm{IF}(x; F) = \frac{\psi(x)}{\int_{-\infty}^{\infty} \psi'(\xi)dF(\xi)}, \tag{9}$$

which here means a scaled version of the function $xe^{-\alpha|x|}$ (Fig. 2).

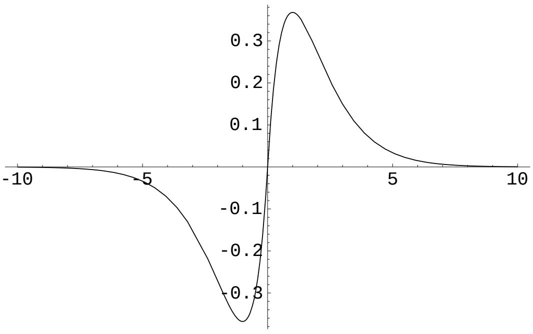
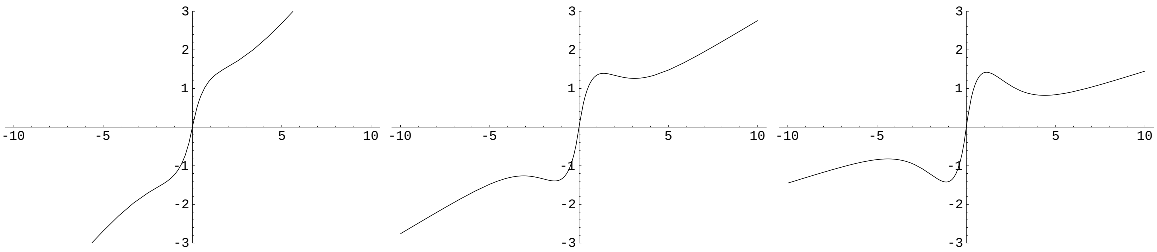


Figure 2: Graph of the function $f(x) = xe^{-\alpha|x|}$, with $\alpha = 1$.

2.1. Robustness properties

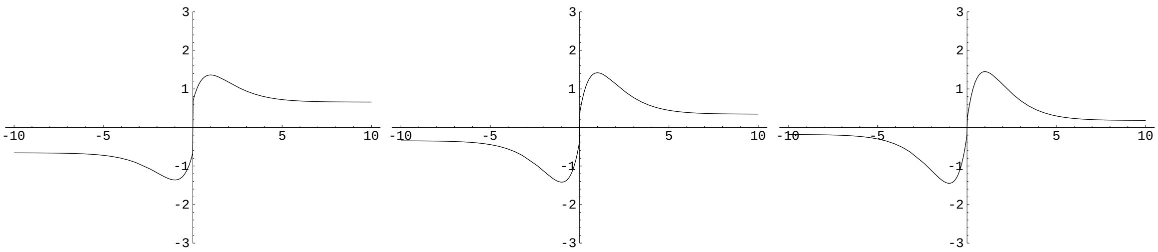
The fact that $\mathrm{IF}_T(x; F)$, the influence function of the estimator used for obtaining the initial reference value, occurs in Eq. 4, is of interest. It shows the significance of the choice of this estimator for the robustness of the RASF in the first rounds (see Fig. 3). The difference



(a) Mean as the initial reference value, 1st round.

(b) Mean as the initial reference value, 2nd round.

(c) Mean as the initial reference value, 3rd round.



(d) Median as the initial reference value, 1st round.

(e) Median as the initial reference value, 2nd round.

(f) Median as the initial reference value, 3rd round.

Figure 3: Influence functions of the first three rounds of the RASF with different initial reference values assuming the standard normal distribution, $\alpha = 1$.

in the speed of convergence of the influence function immediately affects the robustness of the initial stages of the estimator and thus the reliability of the first intermediate outputs.

From Fig. 3 we see that with the median as the initial choice, the influence function approaches $y = xe^{-\alpha|x|}$ (scaled) much faster, inheriting at the same time the robustness of the median, whereas with the mean as the initial choice, the effect of the influence function of the mean is still quite strong in the first rounds, indicating greater non-robustness for more distinct outliers (cf. Fig. 1).

The influence function, when it exists, has a finite gross-error sensitivity and an infinite rejection point. The asymptotic variance of the RASF (with $\alpha = 1$) for the standard normal distribution is $V_{\text{RASF}}(N(0,1)) \approx 1.38$, while $V_{\text{mean}}(N(0,1)) = 1$ and $V_{\text{med}}(N(0,1)) \approx 1.57$. Thus, in this sense the RASF is slightly better than the median but worse than the mean. For Laplace distribution $L(0,1)$ the asymptotic variances are $V_{\text{RASF}}(L(0,1)) \approx 0.60$, $V_{\text{mean}}(L(0,1)) = 1$, and $V_{\text{med}}(L(0,1)) = 1/2$. Again, the RASF is between the mean and the median, closer to the median.

2.2. Convergence

We have obtained a recursive formula for the influence function of the RASF, and, on the other hand, we know that the RASF converges to a limit. The next question is the convergence of the influence function sequence. In other words, does a single sample contribute to the output in a controlled way, that is, is its effect bounded, indicated by the convergence of the influence function sequence to a limit? Moreover, if the sequence converges, what is the speed of its convergence?

We can write the difference of two consecutive influence functions as follows using Eq. 6 for $\text{IF}_{\text{k-1}}(x; F)$:

$$\begin{aligned} &\text{IF}_{\text{k}}(x;F) - \text{IF}_{\text{k-1}}(x;F) = \\ &\quad \tfrac{b}{c}(\text{IF}_{\text{k-1}}(x;F) - \text{IF}_{\text{k-2}}(x;F)). \end{aligned} \tag{10}$$

So, the influence function sequence can be said to converge linearly when $b < c$, in which case the influence function of the RASF is the limit $\frac{xe^{-\alpha|x|}}{c-b}$. The quotient b/c determines the speed of convergence, in that the smaller is b/c (it is positive in all cases), the faster the sequence converges. Likewise, the speed of divergence grows with the quotient b/c. This knowledge is useful if we have fixed beforehand the number of iterations that we are going to use. This situation is very likely, since the computation complexity of the RASF is relatively high, as noted in [3].

The value of b/c depends only on α and the distribution F. In other words, for a given distribution, the weighting parameter α decides the amount of the robustness measurable by the influence function. This means that we can control the robustness of the RASF by setting α suitably. In [3], it was observed that the performance of the RASF in the MAE (Mean Absolute Error) sense depends very heavily on the chosen value for α when the first few iterations are studied.

Now we show that there exist heavy-tailed distributions for which b/c is greater than 1. Let us assume the following distribution:

$$f(x) = 0.5(\delta(x - \frac{1}{\alpha} - 0.001) + \delta(x + \frac{1}{\alpha} + 0.001)), \quad (11)$$

with α given; $\delta(x)$ is the Dirac delta function. To alleviate the non-uniqueness of the median, the mean was used for plotting the influence functions for a first few iterations. Now, with $\alpha = 1$, the influence function sequence diverges for the distribution of Eq. 11 and converges for the standard normal distribution. Fig. 4 shows the diverging case, where the sequence starts with the curve that most closely resembles the line $y = x$ and starts to change towards the general shape of the curve $y = xe^{-|x|}$. However, at the same time the absolute values of the functions of the sequence become arbitrarily large, indicating increasing non-robustness. A converging case is presented in Fig. 5, where the functions approach the curve $y = \frac{xe^{-|x|}}{0.25}$.

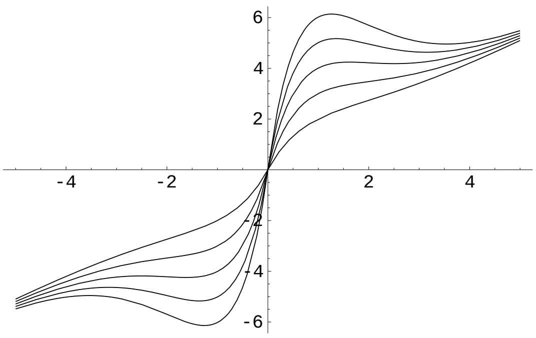


Figure 4: Diverging influence function sequence with the distribution of Eq. 11, $\alpha = 1$. The sequence starts with the curve which is closest to line $y = x$.

In the case of Fig. 4, the divergence means that the estimator becomes more and more non-robust as the number of iterations grow. However, the RASF converges, with the output approaching the point of the impulse on the same side where the deviation from the assumed distribution occurred, in which case the deviation causes a large bias on the output.

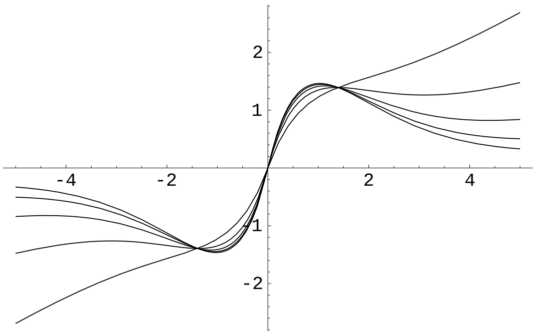


Figure 5: Converging influence function sequence with the standard normal distribution, $\alpha = 1$. The sequence starts with the curve which is closest to line $y = x$.

3. CONCLUSIONS

An influence function was derived for each iteration step of the RASF, and the existence of the influence function for the whole estimator was studied based on the convergence of the sequence of the derived influence functions. From the influence function, it is possible to obtain measures for the robustness of each iteration output and of the final output. A simple test was given for finding out whether the non-robustness of the RASF in the influence function sense stays within bounds or grows arbitrarily large, and examples were given of both situations. The role of the initial filter was shown to be important for the robustness of intermediate results during the first iterations.

4. REFERENCES

[1] F. R. Hampel, E. M. Ronchetti, P. J. Rousseeuw, and W. A. Stahel, *Robust Statistics. The Approach Based on Influence Functions*, Wiley, New York, 1986.

[2] P. J. Huber, *Robust Statistics*, Wiley, New York, 1981.

[3] H. Huttunen, P. Kuosmanen, and J. Astola, "Optimal RASF Filtering," in Nonlinear Image Processing VII, Edward R. Dougherty, Jaakko T. Astola, Harold G. Longbotham, Editors, Proc. SPIE 2662, pp. 154-165, 1996.

[4] P. Kuosmanen and J. Astola, "Connection Between the RASF and M-Estimators," unpublished manuscript.

[5] C. Sun, P. Rabinowitz, "Recursive Approaching Signal Filter," *IEEE Signal Processing Letters,* Vol. 2, No. 5, May 1995.

APPENDIX

The definition of the influence function can also be given as

$$\mathrm{IF}(x;F) = \frac{\partial}{\partial\lambda}\left\{A(F+\lambda(\Delta_x - F))\right\}_{\lambda=0}, \tag{12}$$

where $A(F)$ is the estimator in question expressed in the form of a functional, assuming distribution F.

The functional for the first round of the RASF for distribution F can be expressed as

$$A(F) = \frac{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi dF(\xi)}{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF(\xi)}, \tag{13}$$

so that the influence function for the first round can be derived as follows (the notation $dF(\xi)$ is replaced with dF for simplification):

$$\begin{aligned}
\mathrm{IF}_1(x;F) &= \frac{\partial}{\partial\lambda}\left\{A(F+\lambda(\Delta_x - F))\right\}_{\lambda=0} = \frac{\partial}{\partial\lambda}\left\{\frac{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F+\lambda(\Delta_x - F))|}\xi d(F+\lambda(\Delta_x - F))}{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F+\lambda(\Delta_x - F))|} d(F+\lambda(\Delta_x - F))}\right\}_{\lambda=0} \\
&= \left(\frac{\partial}{\partial\lambda}\left\{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F+\lambda(\Delta_x - F))|}\xi dF\right\}_{\lambda=0} + \int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi d(\Delta_x - F)\right)\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF \\
&\quad \cdot \left(\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF\right)^{-2} - \\
&\quad \left(\frac{\partial}{\partial\lambda}\left\{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F+\lambda(\Delta_x - F))|} dF\right\}_{\lambda=0} + \int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} d(\Delta_x - F)\right)\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi dF \\
&\quad \cdot \left(\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF\right)^{-2} \\
&= \left(\int_{-\infty}^{\infty} \frac{\partial}{\partial\lambda}\left\{e^{-\alpha|\xi - T(F+\lambda(\Delta_x - F))|}\right\}_{\lambda=0}\xi dF + \int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi d\Delta_x - \int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi dF\right) \\
&\quad \cdot \left(\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF\right)^{-1} \\
&= \frac{x e^{-\alpha|x - T(F)|} + \int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\cdot(-\alpha)\frac{\partial}{\partial\lambda}\left\{|\xi - T(F+\lambda(\Delta_x - F))|\right\}_{\lambda=0}\xi dF}{\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|} dF} \\
&= \frac{1}{\int_{-\infty}^{\infty} e^{-\alpha|\xi|} dF}\left(x e^{-\alpha|x|} - \int_{-\infty}^{\infty} \alpha\xi e^{-\alpha|\xi|}\cdot\left\{\begin{array}{ll} \mathrm{IF}_T(x;F), & \xi \le 0 \\ -\mathrm{IF}_T(x;F), & \xi \ge 0 \end{array}\right\} dF\right) \\
&= \frac{x e^{-\alpha|x|} + \int_{-\infty}^{\infty} \alpha\xi e^{-\alpha|\xi|}\mathrm{sign}(\xi)\mathrm{IF}_T(x;F) dF}{\int_{-\infty}^{\infty} e^{-\alpha|\xi|} dF} \\
&= \frac{x e^{-\alpha|x|} + \mathrm{IF}_T(x;F)\int_{-\infty}^{\infty} \alpha|\xi| e^{-\alpha|\xi|} dF}{\int_{-\infty}^{\infty} e^{-\alpha|\xi|} dF}.
\end{aligned} \tag{14}$$

In the derivation we make use of the fact that, with the assumptions stated in Section 2, $\int_{-\infty}^{\infty} e^{-\alpha|\xi - T(F)|}\xi dF = 0$. Derivation and integration are interchanged keeping the usual precautions in mind.